Centre for Communications Research

PDA-BCJR Algorithm for Factorial Hidden Markov Models with Application to MIMO Equalisation

Robert J. Piechocki, Christophe Andrieu, Magnus Sandell and Joe McGeehan

Centre for Communications Research

University of Bristol, Bristol,

Toshiba, TRL Labs, Bristol

University of Bristol

TOSHIBA

ccr centre for communications *research*

Outline of the talk

- MIMO communications problem
- Graphical Models – introduction, application, types, problems in communications, inference algorithms
- proposed PDA-BCJR algorithm
- Numerical results
- Conclusions

Probabilistic Graphical Models - PGM

- **PGMs apply to decision making and/or estimation in the presence of uncertainty**

- **PGMs represent families of probability distribution functions**

- **PGMs do not provide solutions on its own, but can provide:**
  - new insight into existing models
  - motivation for new models
  - suggest construction of new algorithms
  - unified view of problems from smilingly different disciplines of science

Where they are applied and types

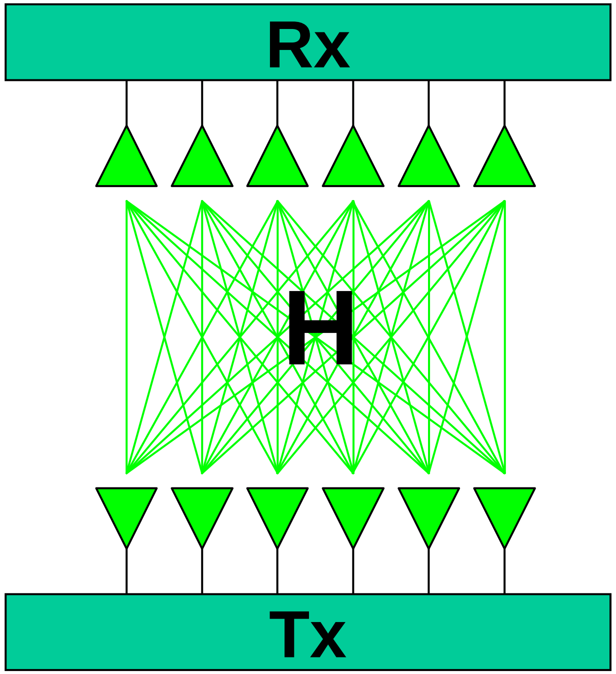
- **Applied (and often independently developed) in:**
  - Bio-informatics (Bio-statistics),
  - Machine Learning (neural nets)
  - Speech processing and image processing
  - Communications, Information retrieval
  - Forensic science
  - and many more

- **There are three most common types of graphs:**
  - Directed acyclic graphs (DAG),
  - Undirected graphs (UG)
  - Factor graphs (FG)

Algorithms for inference in PGMs

- **Exact:**
  - Kalman Filter/Smoother
  - Forward-Backward
  - Sum-product
  - Junction-Tree algorithm (supersedes the above)
- **Monte Carlo (Sampling)**
  - Direct Sampling, Importance sampling
  - MCMC: Gibbs, Metropolis-Hastings
  - Sequential: Particle Filters/Smoothers
- **Deterministic Approximations**
  - Variational Approximation, EM and its variants
  - PDA, Expectation Propagation, GPB, etc..

Considered problem – MIMO communications



It is so called wideband system so the channels are modelled as multi-dimensional FIR filters:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_{1,1}^T & \cdots & \mathbf{h}_{1,N}^T \\ \vdots & \ddots & \vdots \\ \mathbf{h}_{M,1}^T & \cdots & \mathbf{h}_{M,N}^T \end{bmatrix}$$

$$\mathbf{h}_{m,n} = \left(h_{m,n}^{(\tau=0)}, h_{m,n}^{(\tau=1)}, \ldots, h_{m,n}^{(\tau=L-1)} \right)^T$$

Generating model

$$\mathbf{y}^{(t)} = \mathbf{H}\mathbf{x}^{(t)} + \mathbf{n}^{(t)}$$

$$\mathbf{n}^{(t)} \sim \mathcal{N}\left(0, \sigma_n^2 \mathbf{I}\right)$$

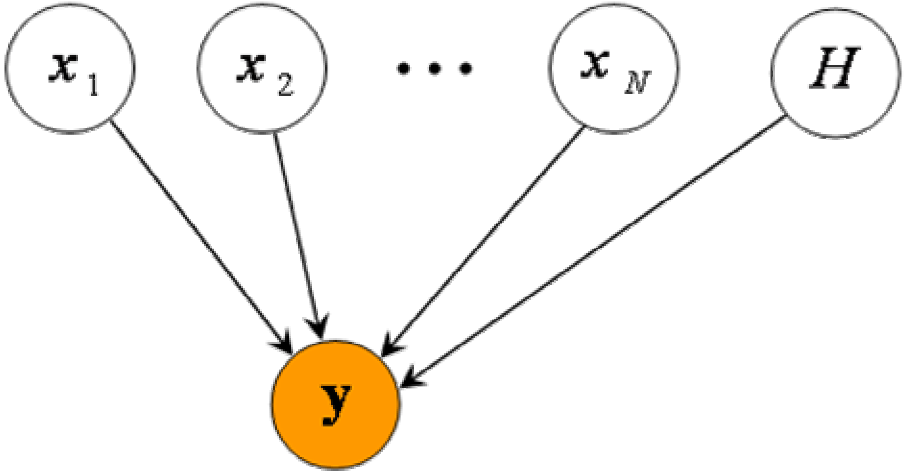
centre for communications research

University of BRISTOL

TOSHIBA

DAG Factorial Model

Communications in non-orthogonal channels: CDMA, MIMO



$$\mathbf{y} = \sum_i \mathbf{h} x_i + \mathbf{n}$$

In our case, **x** – are discrete variables and a marginal of **y** is a mixture Gaussian

$$f(\mathbf{y}, x_1, \ldots, x_N, \mathbf{H}) = f(\mathbf{y} \mid x_1, \ldots, x_N, \mathbf{H})\, f(\mathbf{H}) \prod_{i=1}^{N} f(x_i)$$

Task:

$$f(x_i \mid \mathbf{y}) = \int_{\mathbf{H}} \sum_{x_{-i}} f(x_i, x_{-i}, \mathbf{H} \mid \mathbf{y})\, d\mathbf{H} =$$

$$\int_{\mathbf{H}} \sum_{x_{-i}} f(x_i \mid x_{-i}, \mathbf{H}, \mathbf{y})\, f(x_{-i} \mid \mathbf{H}, \mathbf{y})\, f(\mathbf{H} \mid \mathbf{y})\, d\mathbf{H}$$

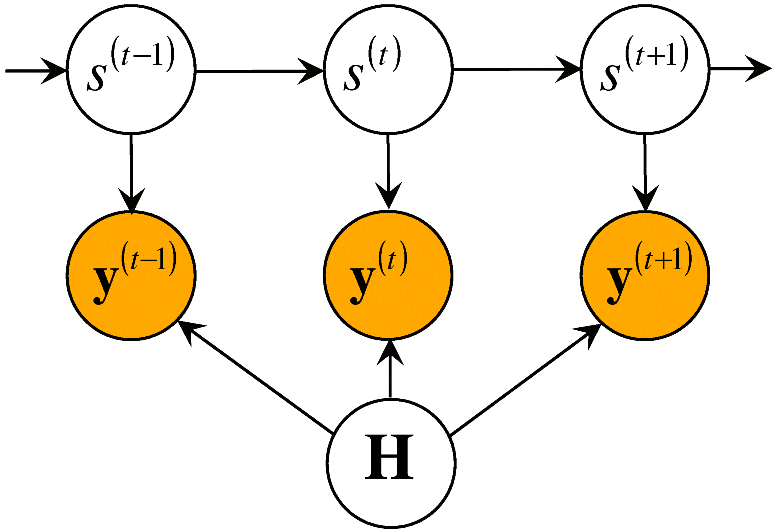
centre for communications research

University of BRISTOL

TOSHIBA

DAG Hidden Markov Model

Received signal in channels with memory, convolutionally encoded signal (binary codes, space-time trellis codes)



$$f\left(\mathbf{y}^{(1)},\ldots,\mathbf{y}^{(T)},s^{(1)},\ldots,s^{(T)},\mathbf{H}\right)=$$
$$f(\mathbf{H})\prod_{t=1}^{T} f\left(\mathbf{y}^{(t)}\middle|s^{(t)},\mathbf{H}\right) f\left(s^{(t)}\middle|s^{(t-1)}\right)$$

States: $s^{(t)} \equiv \left\{x^{(t)},x^{(t-1)},\ldots,x^{(t-L+1)}\right\}$

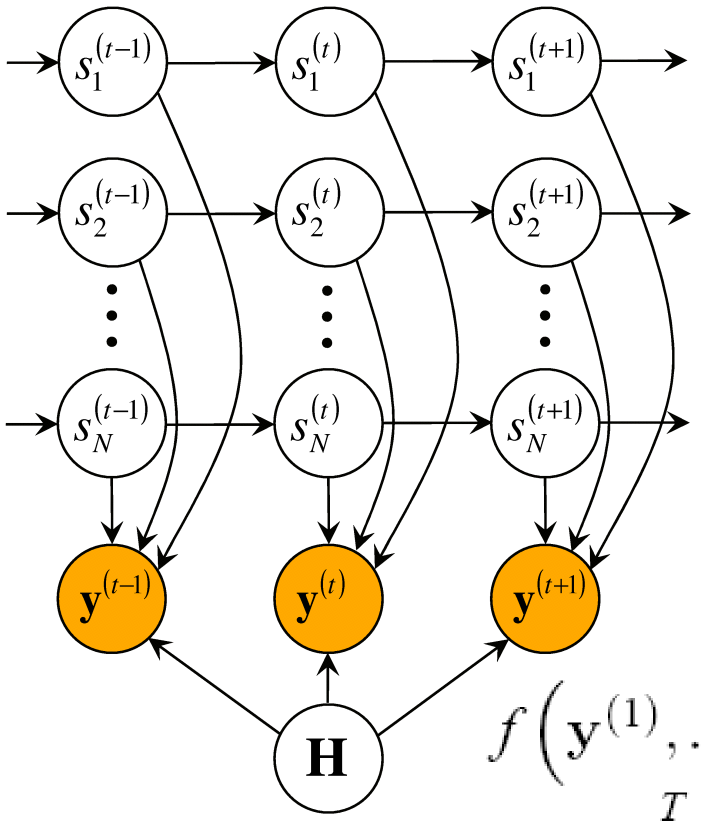
ccr centre for communications research

University of BRISTOL

TOSHIBA

DAG Factorial Hidden Markov Model

Received multiuser (or/and multi-antenna i.e. MIMO) signal in channels with memory



Overall FHMM model arises by replacing the single random variables with HMM

$$f\left(\mathbf{y}^{(1)},\ldots,\mathbf{y}^{(T)},s_1^{(1)},\ldots,s_1^{(T)},\ldots,s_N^{(1)},\ldots,s_N^{(T)},\mathbf{H}\right) =$$
$$f(\mathbf{H})\prod_{t=1}^{T}\left[f\left(\mathbf{y}^{(t)}\middle|s_1^{(t)},\ldots s_N^{(t)},\mathbf{H}\right)\prod_{n=1}^{N}f\left(s_n^{(t)}\middle|s_n^{(t-1)}\right)\right]$$

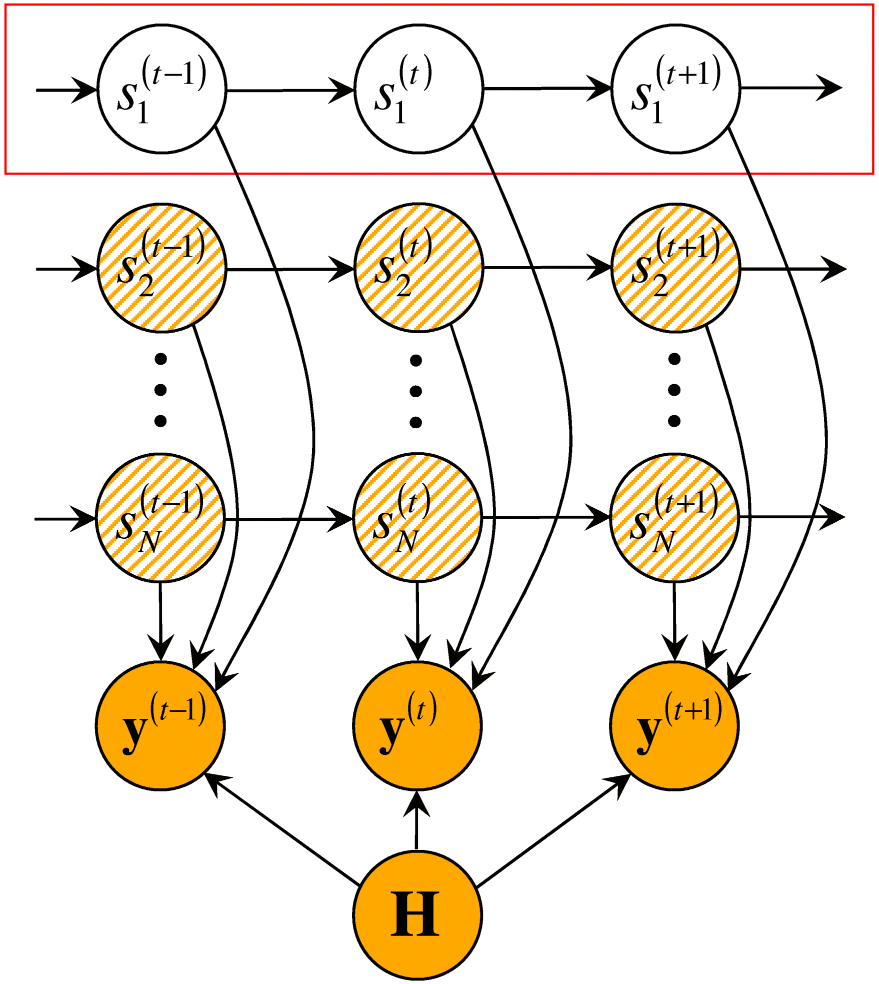
centre for communications research

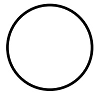
University of BRISTOL

TOSHIBA

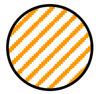
PDA – BCJR Algorithm

Iterations on the chains

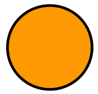




Latent variable (uncertainly fully accounted for)



Latent variable (uncertainly partially accounted for via Gaussian approximation)



Observed variable

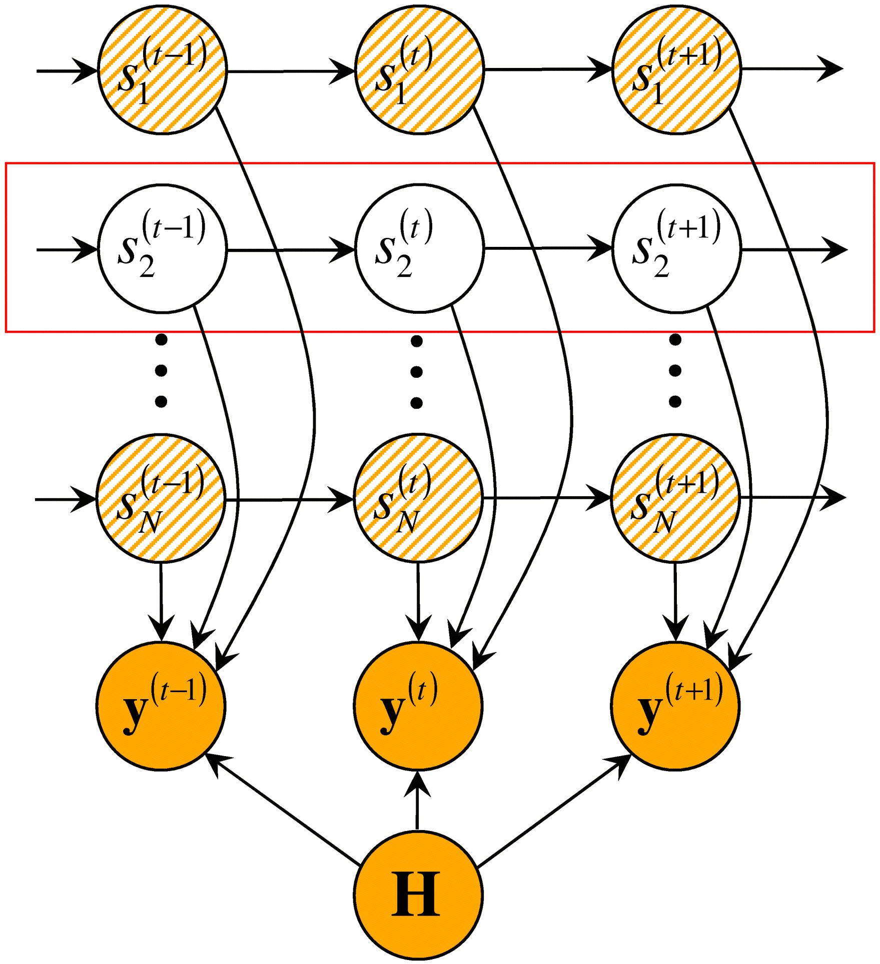
centre for communications research

University of BRISTOL

TOSHIBA

PDA – BCJR Algorithm

Iterations on the chains



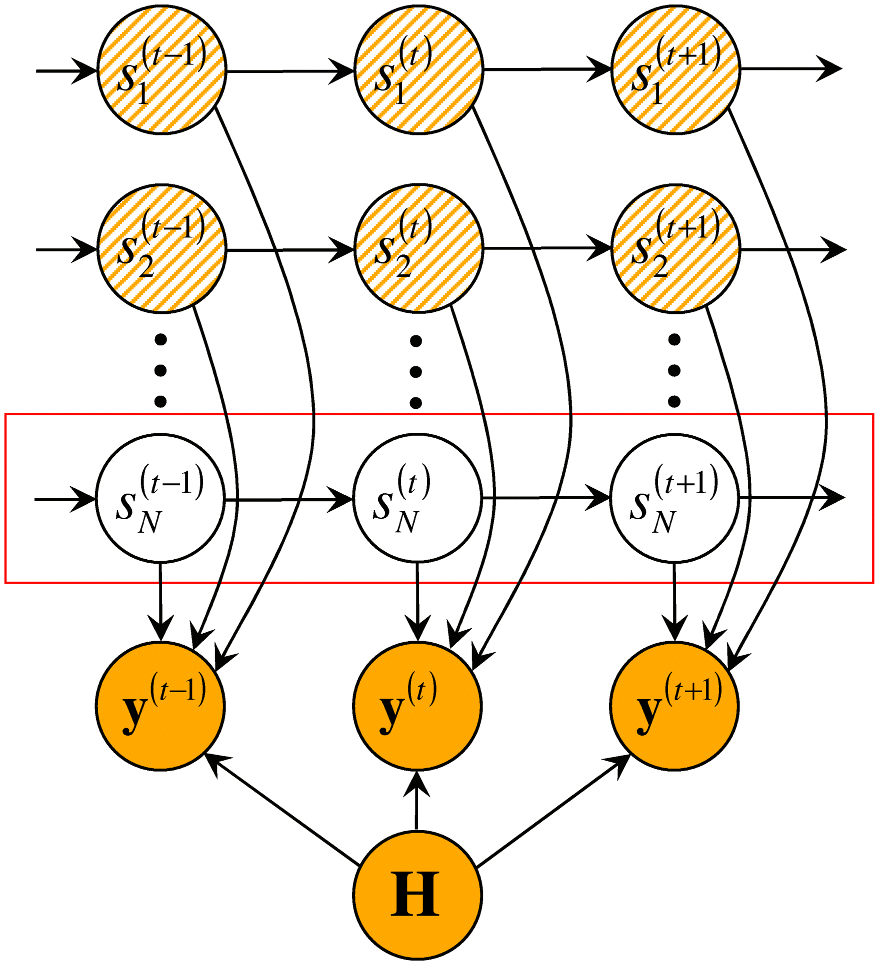
Latent variable (uncertainly fully accounted for)

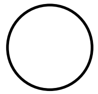
Latent variable (uncertainly partially accounted for via Gaussian approximation)

Observed variable

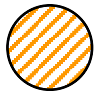
PDA – BCJR Algorithm

Iterations on the chains

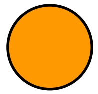




Latent variable (uncertainly fully accounted for)

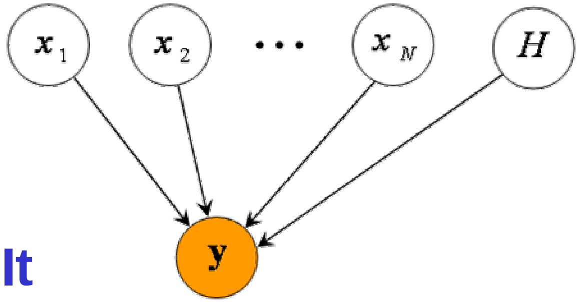


Latent variable (uncertainly partially accounted for via Gaussian approximation)

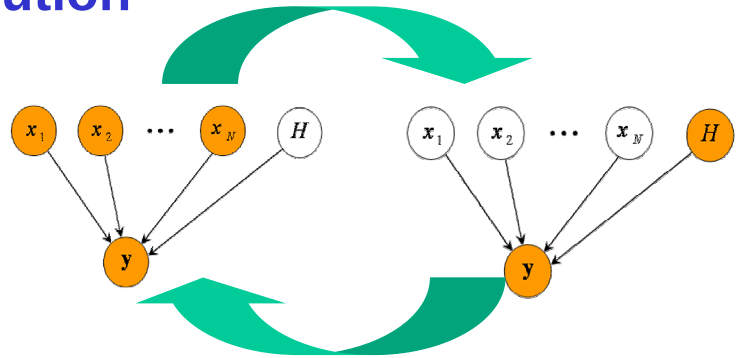


Observed variable

EM – accounting for channel uncertainty



EM solves the "chicken and egg" problem. It is useful where the underlying distribution has a form:



$$f\left(\mathbf{Y}\middle|\theta\right)=\int_X f\left(\mathbf{Y},X\middle|\theta\right)$$

"Chicken and egg"

Set up as estimation (ML or MAP) of **H**, where **X** is the missing data. I.e. we are settling for the expectation of the latent data, rather than the data itself

$$Q\left(\mathbf{H},\mathbf{H}^{(i)}\right)=\underset{\mathbf{x}|\mathbf{H}^{(i)},\mathbf{y}}{E}\left\{\log\left(f\left(\mathbf{H},\mathbf{x}\middle|\mathbf{y}\right)\right)\middle|\mathbf{H}^{(i)},\mathbf{y}\right\}$$

ccr centre for communications research

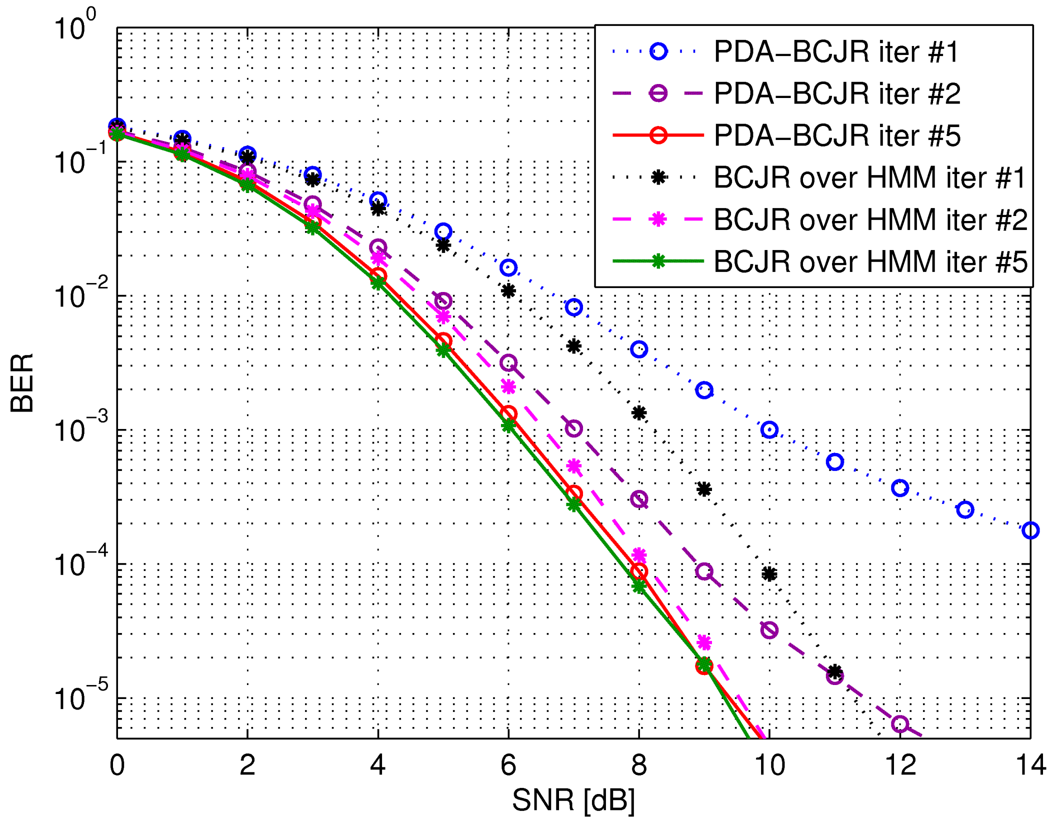
University of BRISTOL

TOSHIBA

PDA-BCJR results I

- PDA-BCJR
- Structured Variational
- Exact (BCJR over HMM)

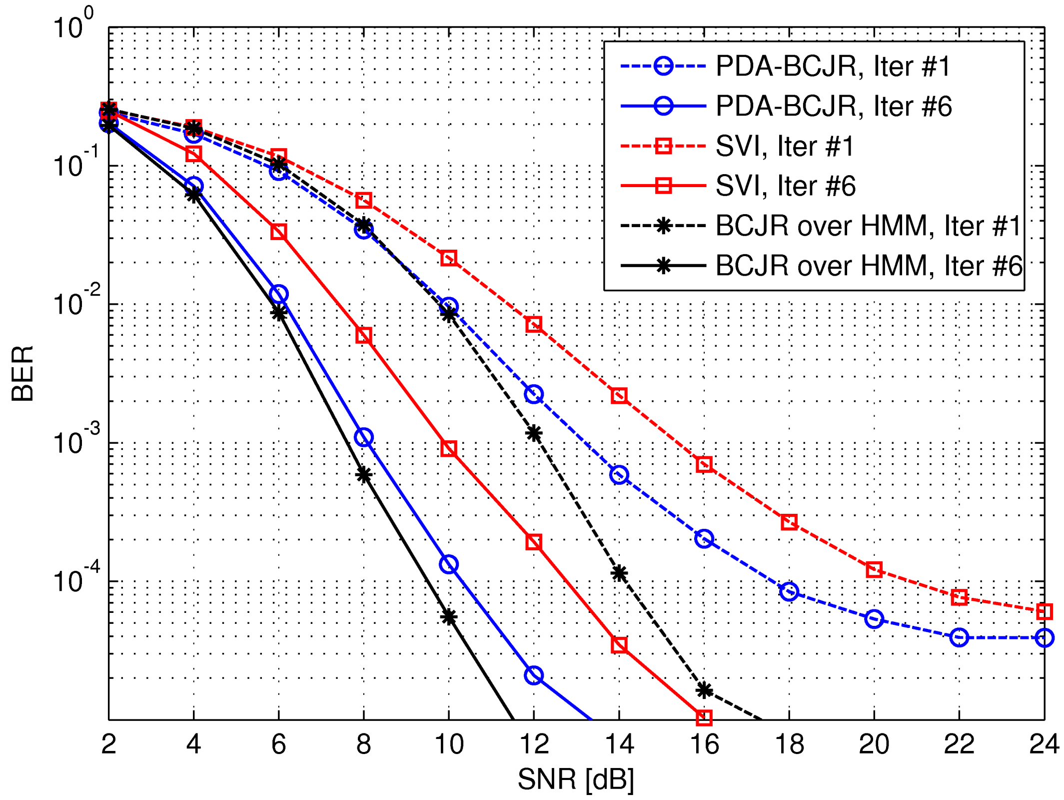
MIMO System with NT=NR=3 antennas, BPSK, 3 tap channel, Channels perfectly known



PDA-BCJR results II – "semiblind setting"

- PDA-BCJR EM
- Structured Variational EM
- BCJR over HMM EM

MIMO System with NT=NR=3 antennas, BPSK, 3 tap channel, first 18 symbols known.



ccr centre for communications research

University of BRISTOL

TOSHIBA

Conclusions

- A generalisation to PDA has been proposed – iterations on entire HMMs
- Particularly suitable in conjunction with EM
- Further degrees of approximation are possible (e.g. within the chains)
- Another application: speech recognition

Additional slides I

Table 1: The PDA-BCJR algorithm summary

"'Outer recursions"'
for $I = 1 : I_{PDAmax}$
"'PDA recursions"'
for $n = 1 : N$
$\mu_{n,in} = \sum_{j=1, j\neq n}^{N} \mu_{j,out}; \quad \mathbf{\Sigma}_{n,in} = \sum_{j=1, j\neq n}^{N} \mathbf{\Sigma}_{j,out}$
$\left[f\left(x_n^{(t),t=1:T} \mid \mathbf{y}\right), \mu_{n,out}, \mathbf{\Sigma}_{n,out} \right] = BCJR^* \left(\mathbf{y}, \hat{\mathbf{H}}_{:,n}, f\left(x_n^{(t),t=1:T}\right), \mu_{n,in}, \mathbf{\Sigma}_{n,in} \right)$
end, end

$$f\left(s', s \mid \mathbf{y}\right) \propto \underbrace{f\left(s', \mathbf{y}^{(1:t-1)}\right)}_{\alpha^{(t)}} \underbrace{f\left(s \mid s'\right) f\left(\mathbf{y}^{(t)} \mid s', s\right)}_{\gamma^{(t)}} \underbrace{f\left(\mathbf{y}^{(t+1:T)} \mid s\right)}_{\beta^{(t)}}$$

$$\gamma^{*(t)}\left(s', s\right) \propto \exp\left(-\left(\mathbf{z}_n^{(t)}\right)^H \left(\mathbf{\Sigma}_{n,in}^{(t)}\right)^{-1} \left(\mathbf{z}_n^{(t)}\right) \right) f\left(x_n\right)$$

Where:

$$\mathbf{z}_n^{(t)} = \mathbf{y}^{(t)} - \hat{\mathbf{H}}_{(:,n)} \mathbf{x}_n - \mu_{n,in}^{(t)}$$

centre for communications research

University of BRISTOL

TOSHIBA

Additional slides II

The output moments are calculated as

$$\mu_{n,out}^{(t)} = E\left\{\hat{\mathbf{H}}_{:,n}\mathbf{x}_n\right\} = \sum_{j=1}^{S}\sum_{k=1}^{K}\hat{\mathbf{H}}_{:,n}\mathbf{x}_n f\left(s', s\,|\mathbf{y}\right)$$

$$\begin{aligned}\boldsymbol{\Sigma}_{n,out}^{(t)} &= E\left\{\left(\hat{\mathbf{H}}_{:,n}\mathbf{x}_n - \mu_{n,out}^{(t)}\right)\left(\hat{\mathbf{H}}_{:,n}\mathbf{x}_n - \mu_{n,out}^{(t)}\right)^H\right\} \\ &= \sum_{j=1}^{S}\sum_{k=1}^{K}\left(\hat{\mathbf{H}}_{:,n}\mathbf{x}_n - \mu_{n,out}^{(t)}\right)\left(\hat{\mathbf{H}}_{:,n}\mathbf{x}_n - \mu_{n,out}^{(t)}\right)^H f\left(s', s\,|\mathbf{y}\right)\end{aligned}$$

ccr centre for communications research

University of BRISTOL

TOSHIBA

Additional slides III

Connections between Variational Inference and PDA

$$f_{PDA}(x_i = a_\omega | \mathbf{y}) \overset{appox}{\propto} \exp\left((\psi - \mathbf{h}_i a_\omega)^T \mathbf{M}^{-1} (\psi - \mathbf{h}_i a_\omega) \right)$$

where:

$$\mathbf{M} = \frac{\sigma^2}{2}\mathbf{I} + \sum_{j=1, j \neq i}^{N} \mathbf{h}_j^T \mathbf{h}_j Var\{x_j\}$$

The relation is:

$$\mathbf{M}_{PDA} = \mathbf{M}_{VI} + \sum_{j=1, j \neq i}^{N} \mathbf{h}_j^T \mathbf{h}_j Var\{x_j\}$$

centre for communications research

University of BRISTOL

TOSHIBA